

Part 1: Short Answer Questions

1. Problem Definition

Hypothetical AI Problem:

Predicting student dropout rates in Kenyan universities.

Objectives:

- Identify students at risk of dropping out early
- Enable timely interventions by counselors
- Improve student retention and graduation rates

Stakeholders:

- University administrators
- Academic counselors and support staff

Key Performance Indicator (KPI):

F1-score of the prediction model

2. Data Collection & Preprocessing

Data Sources:

- Academic records (grades, attendance, enrollment history)
- Learning management system logs (e.g., assignment submissions, forum activity)

Potential Bias:

- Students from marginalized backgrounds might have less complete data, causing underrepresentation and biased predictions.

Preprocessing Steps:

1. Handle missing values through imputation
 2. Normalize numerical features like GPA
 3. One-hot encode categorical variables such as program, gender, and region
-

3. Model Development

Chosen Model:

Random Forest Classifier

Justification:

It handles both numeric and categorical features well, is less prone to overfitting, and provides interpretability.

Data Splitting Strategy:

- 70% training
- 15% validation
- 15% testing
(using stratified sampling)

Two Hyperparameters to Tune:

- `n_estimators`: Controls number of decision trees
 - `max_depth`: Limits tree growth to prevent overfitting
-

4. Evaluation & Deployment

Evaluation Metrics:

- F1-score: Balances false positives and false negatives
- ROC AUC: Measures overall classification performance

Concept Drift:

- Occurs when student behavior patterns change over time (e.g., curriculum updates)
- Monitor by tracking model performance regularly and retraining with updated data

Deployment Challenge:

Integrating the model into existing school systems and dashboards